

## 全プローブ計算①最適化順序

### ・ プローブ単位最適化

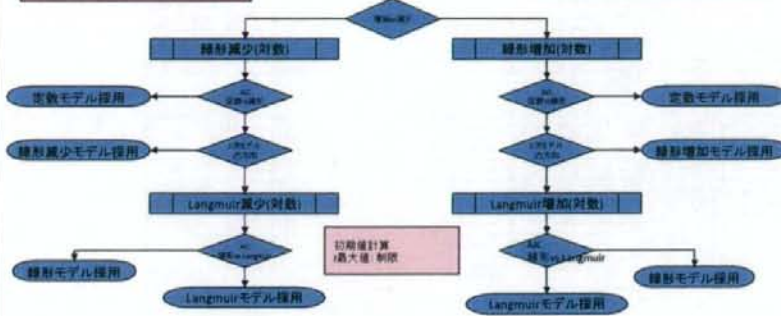
線形(線形)モデルで、傾きがゆるい場合には、定数モデルが選択されるので、Rで計算しない。

定数モデルが線形モデルになる場合には、最終的な定数式は同じなので、計算しない。つまり、下に凸なプローブだけを対象にする。



線形領域の線形モデルで傾き方向と対数領域の線形モデルで逆方向を示すとは考えられない。存在したとしても、定数モデルに落ちると考えられる。

Langmuirは、上に凸な形状である。2次近似モデルで下に凸なモデルとなった場合には、Langmuirモデルは選択してはならない。



## 全プローブ計算①定数モデル(対数領域)

- ・ Teradataを使用し、SQL1文で計算を実施する



尤度

$$L_{x_n}(\mu, \sigma) = \prod_{i=1}^n L_{x_i}(\mu, \sigma) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

対数尤度

$$\begin{aligned} \log L_{x_n}(\mu, \sigma) &= \log \left( \prod_{i=1}^n L_{x_i}(\mu, \sigma) \right) = \log \left( \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \right) \\ &= \sum_{i=1}^n \log \left( \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \right) \\ &= \sum_{i=1}^n \log \left( \frac{1}{\sqrt{2\pi}\sigma} \right) + \sum_{i=1}^n \log \left( e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \right) \\ &= -n \log \sqrt{2\pi} - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \end{aligned}$$

パラメータは、平均と標準偏差の2個

$$AIC = -2\log(L) + 2k$$

$$= 2n \log \sqrt{2\pi} + 2n \log \sigma + \frac{n}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 + 2 \cdot 2$$

## 全プローブ計算①定数モデル(線形領域)

- Teradataを使用し、SQL1文で計算を実施する
  - Reqr\_slope関数を用いる
- 対数尤度・AICは計算しない、傾きの正負を求める

```
select
  probeset_no, probno, pm,
  average(val), reqr_slope(val, lbn_no)
from tbl_cel3 t1
join tbl_exp t2
  on t1.exp_no=t2.exp_no
group by probeset_no, pm, probno
order by probeset_no, pm, probno
```

## 全プローブ計算①二次関数モデル(線形領域)

- Teradataを使用し、SQL1文で計算を実施する
  - 事前に3×3の逆行列を求めておくことにより、行列の積で2次関数モデルを求める
- 対数尤度・AICは計算しない、2次の係数の正負を求める

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ \vdots & \vdots & \vdots \\ 25 & 5 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_n \end{pmatrix}$$

最小2乗法

$$(Kx - v)(Kx - v)$$

$$= (Kx)(Kx) - 2(v)(Kx) + vv$$

$$= x'KKx - 2vKx + vv$$

微分すると

$$'KKx = vK$$

$$x = ('KK)^{-1}vK$$

$$\begin{pmatrix} 0.02381 & -0.14286 & 0.166667 \\ -0.14286 & 0.890476 & -1.1 \\ 0.166667 & -1.1 & 1.533333 \end{pmatrix}$$

```
Select
  probeset_no, pm, probno,
  sum(val) As v1,
  sum(val*lbn_no) As v2,
  sum(val*lbn_no*lbn_no) As v3,
  0.023809524*v1 - 0.142857143*v2 + 0.166666667*v3 As coeff_a,
  -0.142857143*v1 + 0.89047619*v2 - 1.1*v3 As coeff_b,
  0.166666667*v1 - 1.1*v2 + 1.533333333*v3 As coeff_c
From
  tbl_cel3 t1
Join tbl_exp t2
  On
  t1.exp_no=t2.exp_no
Where
  t1.probeset_no=1
Order
  By 1,2,3
```

## 全プローブ計算① Teradataによる線形モデル計算

- Teradataを用いた線形計算をPM補正結果で実施した。
- 線形領域線形モデルの傾きが平均値の2%以上という条件で、減少型133,359プローブ、増加型131,367プローブ、計264,726プローブを実施する。

	Langmuir対象
減少	133,359
増加	131,367
総計	264,726

## 全プローブ計算① Rによる計算方法

- 各プローブに対して、非線形最適化の計算を行う必要がある。
- 次の工夫により、計算速度向上を図る
  - 線形モデル最適化後にLangmuir最適化を行う
    - 100万回の計算回数を25万程度に減らす
  - 1回Rを起動させて、1000probesetsの非線形最適化を実施するように工夫する。
    - 計算途中のRの起動回数を減らし、速度向上に努める

## 全プローブ計算① Langmuirモデル(対数領域)

- Rによる非線形最適化を行う
- 収束結果判定
  - 結果コード=2の場合、kpが大きすぎると失敗
  - Steptol=1.0e-17として、完全不動の状態でしか収束させない
- 初期値
  - $lp$ は、最大値と最小値の差を与える
    - 増加の場合、100:0の値に100:0と25:75の差を加えて、最大値のおおよその値を見つける
  - $Bgp$ は、最小値の半分

## 全プローブ計算① Rスクリプト基本構造

- 線形モデル
  - 入力
    - 識別子:プローブセット番号、PM、プローブ番号
    - データ 15個
      - LBM比率はカラム名に
    - 線形領域計算結果
      - 傾き、切片
  - 初期値
    - 線形領域計算を用いる
  - AIC計算:
    - 定数モデルとAIC判別
    - 標準偏差は小さくなっているはず、なっていないければ、警告
- Langmuirモデル
  - 入力
    - 識別し:プローブセット番号、PM、プローブ番号
    - データ
  - 初期値:
    - データから作成
  - AIC計算
    - 線形モデルとAIC判別
    - 標準偏差は小さくなっているはず、なっていないければ警告

## 全プローブ計算①

### Rスクリプト生成 & R結果ファイル整形プログラム

- Teradataの出力結果を元にRスクリプトを作成するプログラム
  - Visual C++で作成する
- Rスクリプトの実行結果をTeradataへロード可能なファイルにするための整形プログラム
  - AWKで作成する

## 全プローブ計算①計算結果(Cyp7a1)

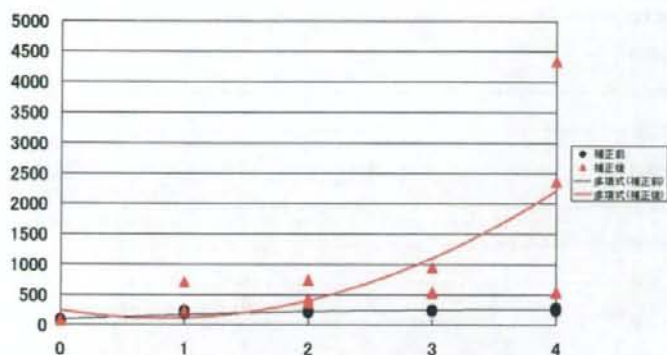
- Cyp7a1に着目してプローブ単位で確認
  - QPCRにてサーフェスの作成がされているので、確認を行いやすいので、このプローブセットを用いた。
- Langmuirモデルが選択されたのは、22プローブ中12プローブ (MM3個、PM9個)

	id	bp	Linear				Langmuir				diff	
			code	iteration	SD	AIC	code	iteration	SD	AIC	SD	AIC
0												
1												
2												
3												
4	225.1261	75.1810	1	9	0.2075	93.28	2	32	0.1281	81.16	0.0783	12.23
5	174.0524	107.8360	1	8	0.1804	85.87	1	28	0.1482	85.28	0.0123	0.39
6												
7												
8	172.5228	38.8222	1	10	0.1394	81.47	1	50	0.1183	78.53	0.0211	2.93
9												
10												
11												
12												
13	380.3265	84.8778	1	8	0.2131	94.18	1	47	0.1881	92.61	0.0240	1.58
14												
15	630.3905	83.2525	1	8	0.1858	90.08	1	42	0.1486	85.58	0.0362	4.50
16	1242.8140	50.8938	1	11	0.1937	91.64	1	66	0.1208	78.12	0.0751	12.52
17	714.4125	105.1452	1	8	0.1278	81.11	1	31	0.1113	78.71	0.0285	4.50
18	631.9981	29.9714	1	13	0.1892	92.18	1	55	0.1295	81.25	0.0687	10.81
19	71309.7788	154.8840	1	10	0.1884	87.30	1	34	0.1180	78.47	0.0513	8.83
20	1138.6972	110.5632	2	18	0.1876	88.99	2	32	0.1033	74.47	0.0843	12.51
21	301.8795	188.5967	1	9	0.1461	82.86	1	34	0.1324	81.91	0.0137	0.95
22	867.8886	188.3824	1	15	0.1564	84.91	1	32	0.1241	79.98	0.0322	4.93



## 全プローブ計算①計算結果(Cyp7a1)

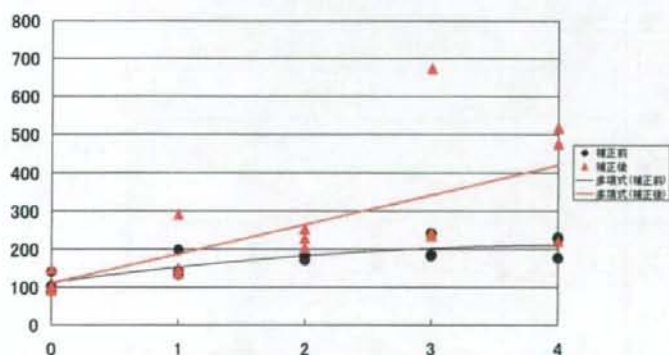
Cyp7a1 MM Probe4



補正後でLiver側で大きな値に変換されている

## 全プローブ計算①計算結果(Cyp7a1)

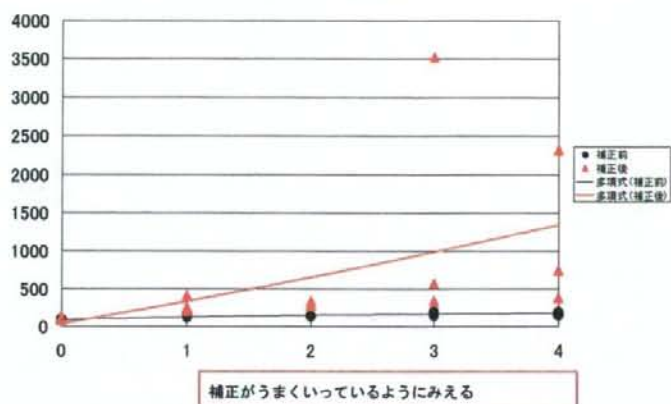
Cyp7a1 MM Probe5



補正がうまくいっているようにみえる

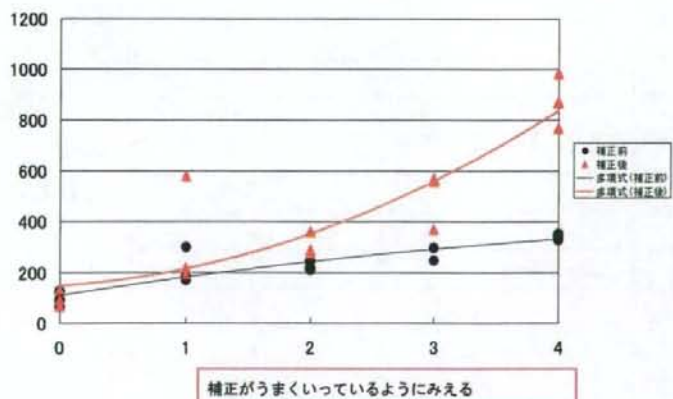
## 全プローブ計算①計算結果(Cyp7a1)

Cyp7a1 MM Probe8



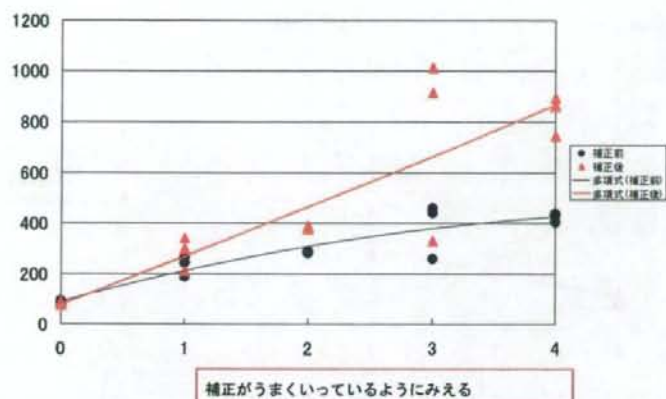
## 全プローブ計算①計算結果(Cyp7a1)

Cyp7a1 PM Probe1



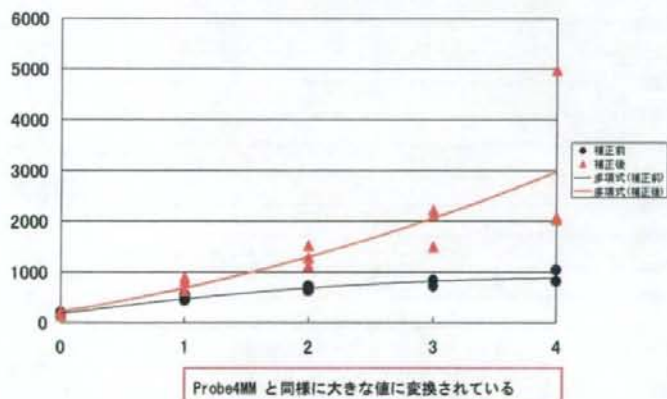
## 全プローブ計算①計算結果(Cyp7a1)

Cyp7a1 PM Probe3



## 全プローブ計算①計算結果(Cyp7a1)

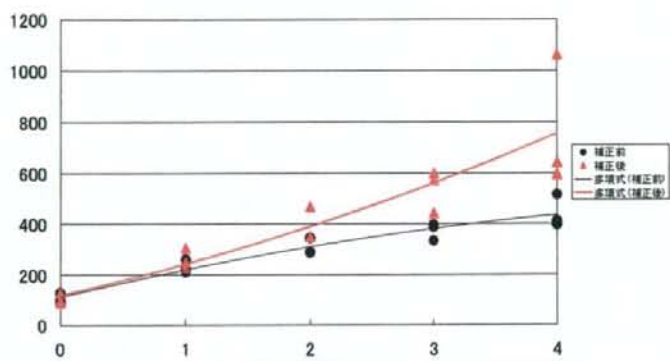
Cyp7a1 PM Probe4





## 全プローブ計算①計算結果(Cyp7a1)

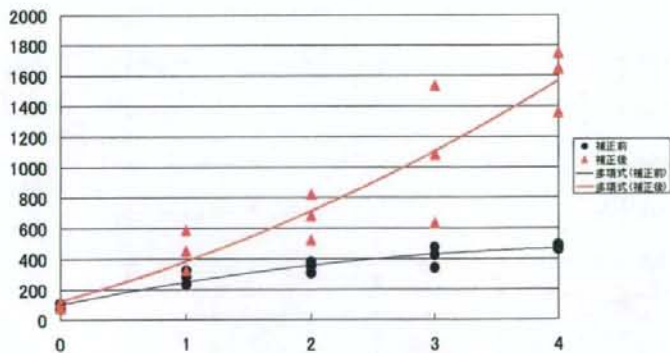
Cyp7a1 PM Probe5



補正がうまくいっているようにみえる

## 全プローブ計算①計算結果(Cyp7a1)

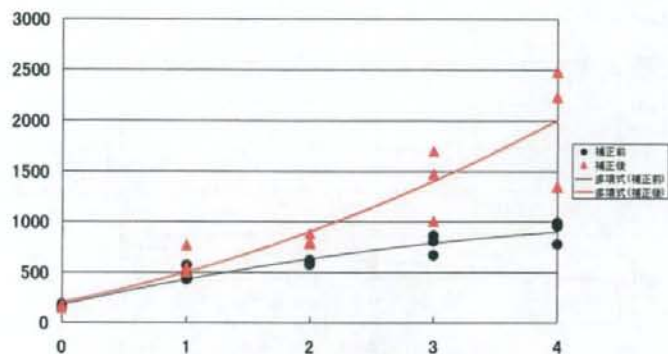
Cyp7a1 PM Probe6



補正がうまくいっているようにみえる

## 全プローブ計算①計算結果(Cyp7a1)

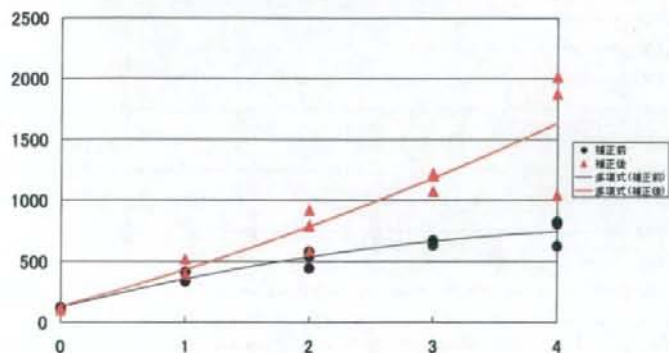
Cyp7a1 PM Probe7



補正がうまくいっているように見える

## 全プローブ計算①計算結果(Cyp7a1)

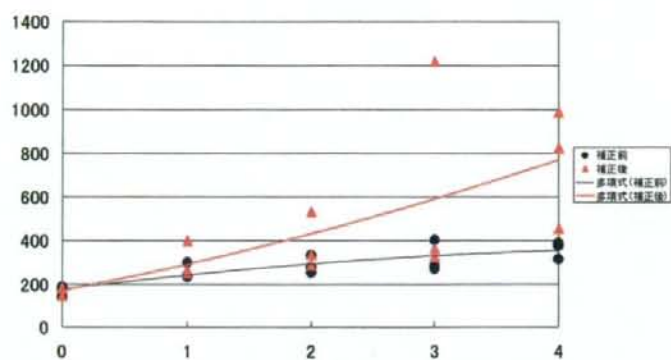
Cyp7a1 PM Probe8



補正がうまくいっているように見える

### 全プローブ計算①計算結果(Cyp7a1)

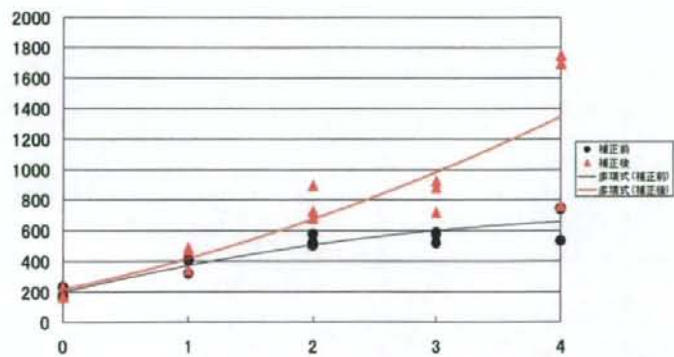
Cyp7a1 PM Probe9



補正がうまくいっているようにみえる

### 全プローブ計算①計算結果(Cyp7a1)

Cyp7a1 PM Probe10



補正がうまくいっているようにみえる

## 全プローブ計算②予備検討

- 全プローブに対して計算を一通り可能であるか確認した。

## 全プローブ②非線形最適化の結果

- Liver側飽和では多くのプローブでLangmuirモデルが採用された。
- Brain側飽和では、Langmuirモデルが採用されたプローブは限られていた。

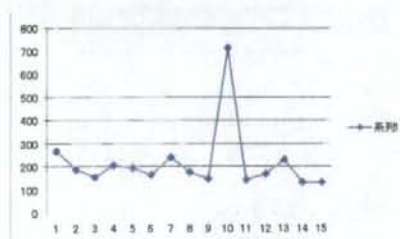
	Liver側飽和	Brain側飽和
計算	87,288	40,299
Langmuir採用	35,161	246
失敗: 収束回数超過	0	1
失敗: $\sigma$ 減少せず	708	1,548

### 全プローブ計算②収束失敗(収束回数超過)のプローブ

- Brain側飽和条件のプローブ28083/mm/7はLangmuirモデルで、収束回数が超過するという結果になった。一実験だけ外れた値を持っているために収束失敗したと考えられる。

### 全プローブ計算②収束失敗( $\sigma$ 減少不十分)のプローブ

- パラメータ数を増加させれば、標準偏差は減少するはずであるが、はずれ値を含むなどの場合にはうまくいかないことも考えられる。このような場合には、Langmuirモデルが採用できなかったものとみなす。新たなLBMで係数決定を行えばLangmuirモデルが採用される可能性が高い。



失敗プローブ例  
4806-pm-9

## 全プローブ計算② Langmuir Model採用個数

- ProbeSet中で何個のProbeがLangmuirモデルを採用しているか
- Brain側飽和

		MM Langmuir Model個数			総計
		1	2	11	
PM Langmuir Model数	3	2			2
	5	2			2
	10		1		1
	19			1	1
総計		4	1	1	6

PM側のLangmuirモデル採用個数が多いProbeSetが補正の影響を受けるはずである  
5個以上と考えると、4ProbeSet

## 全プローブ計算② Langmuir Model採用モデル

- ProbeSet中で何個のProbeでLangmuirモデルを採用しているか
- Liver側飽和

		MM Langmuir Model個数											総計			
		1	2	3	4	5	6	7	8	9	10	11				
PM Langmuir Model数	1	855	97	5												957
	2	204	51	4												259
	3	89	28	6	1	1		1								125
	4	74	28	12	2											116
	5	58	39	22	4	1										122
	6	57	40	22	3	1	2									125
	7	52	46	32	10	6	2	1								149
	8	30	37	52	23	13	3	4				1				163
	9	25	33	40	37	23	20	8	3	3						192
	10	12	22	34	32	38	30	37	21	13	9	2				250
	11	4	7	13	21	43	47	65	81	81	83	105				550
	12						1									1
	15			1												1
	総計		1458	428	243	133	126	105	116	105	97	93	107			3011

PM側のLangmuirモデル採用個数が多いProbeSetが補正の影響を受けるはずである  
5個以上と考えると、1500個程度



### 全プローブ計算③

- 非線形最適化の課題解決を試みて全プローブの補正計算結果を求め、サマライズ(あるプローブセットに属するプローブ群の測定値を取りまとめ、プローブセットとしての測定値を算出すること)後の結果について考察する。

### 全プローブ計算③非線形最適化の課題(Langmuirモデル)

- 収束結果判定
  - 結果コード=2の場合、 $k_p$ が大きすぎると失敗
  - $Stoptol=1.0e-17$ として、完全不動の状態でしか収束させない
- 初期値
  - $lp$ は、最大値と最小値の差を与える
    - 増加の場合、100:0の値に100:0と25:75の差を加えて、最大値のおおよその値を見つける
  - $Bgp$ は、最小値の半分

### 全プローブ計算③非線形最適化の課題(解決テクニック)

- いくつかのプローブにおいては、適切な解に収束しなかった
  - 繰り返しの範囲量を超えた。または収束回数を超えた。
    - 非線形計算のパラメータを変更してやり直すようにプログラムに組み込み、ほとんどのプローブで収束するようにした
  - 繰り返しの最中に無限の数値が現れ、パッチモードのRが中断した
    - メッセージ「nim」により有限でない値が与えられました」
    - 手作業でパラメータ変更実施

補正方法	飽和側	モデル	ProbesetNo	Psi/Min	Probe	対応方法
PM	Liver	Langmuir	10709	MM	6	IOO=18.5, BPO=16
	Liver	Linear	43161	PM	3	BPOを半分に
	Brain	Linear	1680	PM	10	BPOを半分に
	Brain	Linear	35906	PM	5	BPOを半分に
PMMM	Liver	Langmuir	11913	pm	4	IOO=92, kOO=1.11, BPO=11.9
	Brain	Linear	15256	pm	4	BPOを半分に

今後も発生する可能性は高いが、既に来上がったモデルのパラメータを与えることで、回避可能と考える。

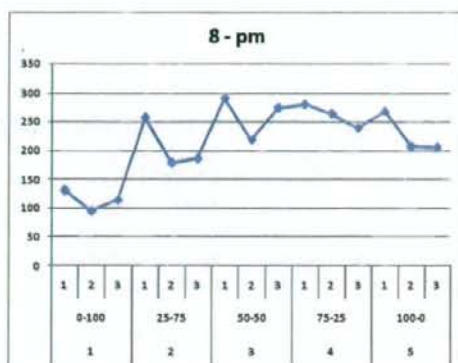
### 全プローブ計算③非線形最適化の結果

- Liver側飽和では多くのプローブでLangmuirモデルが採用された。
- Brain側飽和では、Langmuirモデルが採用されたプローブは限られていた。
- 標準偏差が減少しないという、理論と外れた結果が存在した。
  - 収束先の存在しないモデルに対して最適化を実施した可能性が高い

	Liver側飽和	Brain側飽和
計算	109,879	130,706
Langmuir採用	77,684	9,094
失敗: $\sigma$ 減少せず	127	8

### 全プローブ計算③収束失敗( $\sigma$ 減少不十分)のプローブ

- パラメータ数を増加させれば、誤差は減少するはずであるが、はずれ値を含むなどの場合にはうまくいかないことも考えられる。このような場合には、Langmuirモデルが採用できなかったものとみなす。



例: 2811-8-PM

### 全プローブ計算③ Langmuir Model採用個数(Brain側飽和)

- ProbeSet中で何個のProbeがLangmuirモデルを採用しているか
- Brain側飽和

合計 / Count(*)	n MM	1	2	3	13 ?	総計
1	409	45	5			3063
2	73	4	3			314
3	14					45
4	4	2				9
5	2	2				1
6		2				1
7	1		1			2
8	1	2				3
10		1				1
20					1	1
総計	3194	257	16			3467
	3698	315	25	1		7492

PM側のLangmuirモデル採用個数が多いProbeSetが補正の影響を受けるはずである

### 全プローブ計算③ Langmuir Model採用個数(Liver側飽和)

- Liver側飽和

番付 / Count(n)	n MM													総計		
n PM	1	2	3	4	5	6	7	8	9	10	11	12	13			
1	2639	663	89	13	1										4277	7692
2	780	345	57	11	1										359	2132
3	296	130	41	6	2				1						299	765
4	159	85	41	16	2	2	1								126	432
5	103	55	46	23	10	3	1								84	364
6	93	82	50	22	16	10									53	326
7	78	77	56	47	22	10	4	1							50	345
8	66	67	67	49	40	22	12	5	1						27	356
9	31	57	85	67	47	41	27	14	6	2	1				9	387
10	21	31	50	73	59	73	75	59	37	15	7				7	507
11	3	13	21	42	63	59	117	154	165	190	255				3	1100
12						1										1
13								1						1		2
14																1
15																1
16																1
17																1
18																1
19																1
n	8005	2273	85	10												7026
総計	15262	2470	672	378	263	226	227	234	211	207	274	1	1	1	5871	21407

PM側のLangmuirモデル採用個数が多いProbeSetが補正の影響を受けるはずである

### 全プローブ計算③ Langmuir補正時エラーファイル

- 飽和モデルを用いて次の状態となった場合、エラーファイルに出力する
  - モデル作成LBMの最小最大の範囲を逸脱した
  - モデル作成LBMの $3\sigma$ の範囲を逸脱した
  - Langmuirモデルが採用されているプローブでシグモイド関数の(0,1)の範囲を逸脱した
    - バックグラウンドレベル( $b_{bg}$ )を下回る
    - 飽和レベル( $b_s$ )を上回る
- モデル作成LBMの範囲を逸脱している状態を参考に、追加すべき標準データの作成を考える
- Langmuirモデル範囲の逸脱に関して、何らかの対処が必要

### 全プローブ計算③検証①

- LBMでの線形性チェック
  - 50:50を1に正規化したグラフを作成し、線形性のチェック
  - Langmuir変換を含むプローブセットを全てプロットし、目視で線形性の確認する。値の低いもの以外で異常な値を示すものがないことを確認する

### 全プローブ計算③検証①

- 対象プローブセット:50%:50%において、1000程度の値を持つ

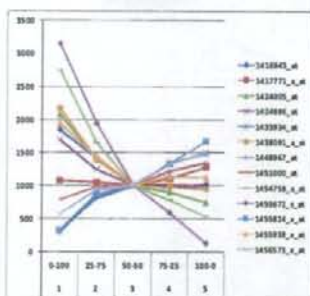
afly_id	Intensity
1424886_at	1008.53
1433934_at	1004.20
1424005_at	1002.17
1417771_a_at	1005.90
1451000_at	1000.37
1456573_x_at	1005.87
1455824_x_at	1006.07
1416943_at	1007.17
1454758_a_at	1007.60
1438091_a_at	1008.33
1448967_at	1003.97
1455939_x_at	1006.53
1455672_s_at	1004.00



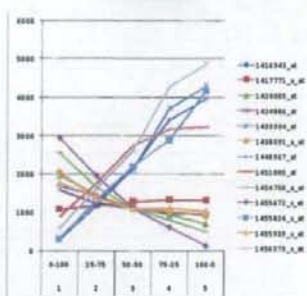
### 全プローブ計算③検証①

- 対象プローブセット: 50%:50%において、1000程度の値を持つ

補正前



補正後

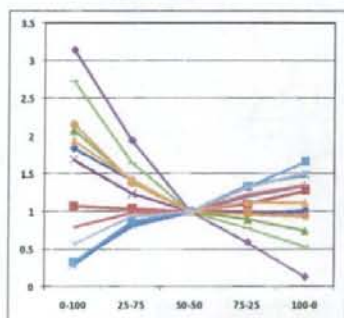


Liver側飽和のプローブは直線に引き延ばされ、高い値で、不安定になっている。

### 全プローブ計算③検証①

- 50:50で1に統一

補正前



補正後

